



No Evidence for Association of Autism with Rare Heterozygous Point Mutations in Contactin-Associated Protein-Like 2 (CNTNAP2), or in Other Contactin-Associated Proteins or Contactins

Citation

Murdoch, J. D., A. R. Gupta, S. J. Sanders, M. F. Walker, J. Keaney, T. V. Fernandez, M. T. Murtha, et al. 2015. "No Evidence for Association of Autism with Rare Heterozygous Point Mutations in Contactin-Associated Protein-Like 2 (CNTNAP2), or in Other Contactin-Associated Proteins or Contactins." PLoS Genetics 11 (1): e1004852. doi:10.1371/journal.pgen.1004852. <http://dx.doi.org/10.1371/journal.pgen.1004852>.

Published Version

doi:10.1371/journal.pgen.1004852

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:13890742>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

RESEARCH ARTICLE

No Evidence for Association of Autism with Rare Heterozygous Point Mutations in Contactin-Associated Protein-Like 2 (*CNTNAP2*), or in Other Contactin-Associated Proteins or Contactins



CrossMark
click for updates

John D. Murdoch^{1*}, Abha R. Gupta^{2,3}, Stephan J. Sanders^{1,4}, Michael F. Walker⁴, John Keaney⁵, Thomas V. Fernandez², Michael T. Murtha², Samuel Anyanwu², Gordon T. Ober², Melanie J. Raubeson², Nicholas M. DiLullo², Natalie Villa⁶, Zainab Abdul Waqar², Catherine Sullivan², Luis Gonzalez², A. Jeremy Willsey^{1,4}, So-Yeon Choe⁷, Benjamin M. Neale^{8,9}, Mark J. Daly^{8,9}, Matthew W. State^{1,2,4*}

1 Yale University Department of Genetics, New Haven, Connecticut, United States of America, **2** Yale University Child Study Center, New Haven, Connecticut, United States of America, **3** Yale University Department of Pediatrics, New Haven, Connecticut, United States of America, **4** University of California, San Francisco, Department of Psychiatry, San Francisco, California, United States of America, **5** Yale College, New Haven, Connecticut, United States of America, **6** UCLA David Geffen School of Medicine, Los Angeles, California, United States of America, **7** UCLA School of Law, Los Angeles, California, United States of America, **8** Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America, **9** Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States of America

* Current address: European Neuroscience Institute, Göttingen, Germany

* matthew.state@ucsf.edu

OPEN ACCESS

Citation: Murdoch JD, Gupta AR, Sanders SJ, Walker MF, Keaney J, Fernandez TV, et al. (2015) No Evidence for Association of Autism with Rare Heterozygous Point Mutations in Contactin-Associated Protein-Like 2 (*CNTNAP2*), or in Other Contactin-Associated Proteins or Contactins. PLoS Genet 11(1): e1004852. doi:10.1371/journal.pgen.1004852

Editor: Jonathan Flint, The Wellcome Trust Centre for Human Genetics, University of Oxford, UNITED KINGDOM

Received: May 28, 2014

Accepted: October 26, 2014

Published: January 26, 2015

Copyright: © 2015 Murdoch et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All raw data are publicly available at the NDAR website (https://ndar.nih.gov/data_from_labs.html) under the heading "Genomic Profiling and Functional Mutation Analysis in Autism Spectrum Disorders" and accession #1895. Approved researchers can obtain the SSC biospecimens described in this study (<http://sfari.org/resources/simons-simplex-collection>) by applying at <https://base.sfari.org> and completing the Researcher

Abstract

Contactins and Contactin-Associated Proteins, and Contactin-Associated Protein-Like 2 (*CNTNAP2*) in particular, have been widely cited as autism risk genes based on findings from homozygosity mapping, molecular cytogenetics, copy number variation analyses, and both common and rare single nucleotide association studies. However, data specifically with regard to the contribution of heterozygous single nucleotide variants (SNVs) have been inconsistent. In an effort to clarify the role of rare point mutations in *CNTNAP2* and related gene families, we have conducted targeted next-generation sequencing and evaluated existing sequence data in cohorts totaling 2704 cases and 2747 controls. We find no evidence for statistically significant association of rare heterozygous mutations in any of the *CNTN* or *CNTNAP* genes, including *CNTNAP2*, placing marked limits on the scale of their plausible contribution to risk.

Distribution Agreement which ensures these are being used for appropriate scientific purposes. The Baylor-Broad data are from the "Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls" study (PMID 23593035), whose authors may be contacted at roeder@stat.cmu.edu and mjdaly@atgu.mgh.harvard.edu.

Funding: This work was supported by grants R01 MH081754 and RC2 MH089956 to MWS, and by K08MH087639 to ARG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Prior genetic studies of autism spectrum disorders (ASD) have demonstrated a role for Contactin-Associated Protein-Like 2 protein (*CNTNAP2*), as well as for other genes that code for Contactin proteins and Contactin-Associated Proteins. While there is strong evidence that the loss of two copies of the gene *CNTNAP2* causes autism and epilepsy, the impact of mutations in only one copy of this gene, or in only one copy of related genes, is less clear. We performed large-scale DNA sequencing on a cohort of over 1000 autism patients and nearly 1000 unaffected controls and did not find significant association at any of 6 genes in the Contactin family and 4 genes in the Contactin-Associated Protein family when looking for rare mutations that are predicted to be disruptive to the protein's function and are present in only one copy of the respective gene. We then combined the data on *CNTNAP2* from our laboratory with *CNTNAP2* data from another research laboratory, and found no significant association of deleterious heterozygous mutations at this gene. Given the paucity of nonsense mutations identified across the combined sample, an assessment of their impact was circumscribed. However, missense heterozygous mutations in *CNTNAP2* and in other Contactins or Contactin-Associated Proteins are not elevated in affected individuals versus controls and, consequently, do not have a marked impact, as a group, on the risk for autism spectrum disorders.

Introduction

Contactin-associated protein-like 2 (*CNTNAP2*; a.k.a *CASPR2*, OMIM 604569, Swiss-Prot Q9UHC6), is widely considered an established autism spectrum disorder (ASD, OMIM 209850) risk gene based on a series of findings including homozygosity mapping of rare recessive mutations [1], cytogenetics [2, 3], deep re-sequencing [2], common variant candidate gene association studies [3, 4], and behavioral phenotyping of mouse models [5]. Moreover, several of the reported human genetic studies have extended the linked or associated phenotype to include language delay [6, 7], intractable epilepsy [1], cortical malformation [1] and intellectual disability [8, 9] (OMIM 610042) and in one instance, hepatomegaly and periventricular leukomalacia [10]. These findings, in aggregate, point to an important role for *CNTNAP2* in the development of the human central nervous system and provide strong evidence that the complete loss of function of this protein leads to autism and epilepsy. However, replication of the genetic association studies with regard to heterozygous variants in ASD^{2–4} has proven elusive. With regard to common variation, three genome wide association studies have not found further significant evidence for previously identified single nucleotide polymorphisms (SNPs) [11–13] at *CNTNAP2* while a fourth found suggestive evidence falling short of genome-wide significance for a nearby SNP [14]. Moreover, a recent targeted study of common alleles at the *CNTNAP2* locus did not replicate prior findings or provide new evidence for significant associations in this interval, after correction for multiple comparisons and the inclusion of a larger follow-up cohort [15]. In addition, to our knowledge, there have been no further published investigations of rare variant mutation burden in *CNTNAP2* beyond the initial report [2] from our laboratory.

In an effort to further explore the evidence for a role for rare heterozygous mutations in *CNTNAP2* in ASD, we set out to re-sequence a new discovery cohort including 1030 individuals affected with ASD as well as 942 psychiatrically unscreened controls, using a combination of micro-emulsion PCR and next generation sequencing. Moreover, we extended our investigation to include the additional genes Contactins 1–6 (OMIM 600016, 190197, 601325, 607280,

607219, 607220, respectively; Swiss-Prot Q12860, Q02246, Q9P232, Q8I WV2, O94779, Q9UQ52, respectively) and Contactin Associated Proteins 1, 3, 4 and 5 (OMIM 602346, 610517, 610518, 610519, respectively; Swiss-Prot P78357, Q9UHC6, Q9BZ76, Q9C0A0, Q8WYK1, respectively), based on a range of evidence for several of these neuronal adhesion molecules in neurodevelopmental phenotypes [16–18].

In our prior analysis of *CNTNAP2*, sequencing of a separate cohort of 635 cases and 942 controls led to the identification of one nominally significant rare variant, I869T, found in three apparently unrelated cases and not in controls, though at the time of publication we were not able to rule out a role for population stratification in this finding [2]. In addition, we observed an approximately 2 fold increase in rare highly conserved (across 10 vertebrate lineages) mutations in cases versus controls, but this fell just short of statistical significance. In the current study, after correction for multiple comparisons, we find no additional support for an association of the I869T variant, which in contrast to our prior study, is no longer predicted to be deleterious to protein function [19, 20]. Moreover, we find no evidence for association of rare heterozygous mutations in general in *CNTNAP2* with the ASD phenotype, either in our new discovery cohort (N = 1030 cases and 942 controls), among an independent ASD sample that had previously undergone exome sequencing (N = 1039 ASD cases and 863 controls), or in a combined mutation burden analysis including these two data sets along with the cohort previously sequenced in our laboratory [2] (N = 2704 cases and 2747 controls). Our ability to assess the impact of rare loss of function mutations was limited by very few observations. However, our findings with regard to rare missense variants was not altered by predicting putative functional versus neutral variants using a range of informatics tools. Finally, we found no significant support for association of rare heterozygous mutations in other Contactins or Contactin-Associated Proteins via targeted re-sequencing of our discovery cohort (1030 cases and 942 controls).

Given the strong evidence for a link between ASD, epilepsy and homozygous loss of *CNTNAP2*, our data do not rule out a role for this gene and associated molecules in the risk for ASD, but place further bounds on the plausible size of the contribution from rare heterozygous missense mutations in these genes.

Results

Burden of rare mutations in cases and controls

In prior work [2] we had noted a non-significant overrepresentation of all rare (<1 observation in 4000 combined [sequenced or TaqMan-genotyped] controls) variants in affected individuals. Consequently, our primary hypothesis was that in an expanded analysis, the burden of rare mutations would be greater in cases than in well-matched controls. After QC steps and exclusion of ancestral outliers, we had a total of 1030 cases and 942 controls and conducted a gene by gene analysis, including variants that were missense, nonsense, or splice site-disrupting, fell below 2% in the high-throughput data, and were filtered against population data as described in Methods below. Variants that were present only once in cases or controls are denoted as “singleton” and are estimated to have an allele frequency approximate to that in our prior analysis (<1 observation/4000)[2]. For *CNTNAP2*, we found no increase in the overall burden of rare mutations (Table 1, Fig. 1) even before correcting for multiple comparisons and in fact more rare mutations in controls than cases (19 in cases, 26 in controls, Fisher exact one-tailed $p = 0.113$). We then evaluated all *CNTN* and *CNTNAP* genes and nominal association for rare heterozygous mutations in *CNTN1* (14 in cases, 5 in controls, Fisher exact one-tailed $p = 0.048$) and *CNTN3* (13 in cases, 3 in controls, Fisher exact one-tailed $p = 0.016$), however, neither results survived experiment-wide correction for multiple comparisons.

Table 1. Overall rare* variant mutation burden: all genes.

Gene	Number of rare mutations in cases (n = 1030)	Number of rare mutations in controls (n = 942)	OR	OR 95% CI lower	OR 95% CI upper	One-tailed Fisher exact p-value
<i>CNTN1</i>	14	5	2.582	0.867	8.227	0.048
<i>CNTN2</i>	14	12	1.068	0.463	2.473	0.514
<i>CNTN3</i>	13	3	4.001	1.065	17.707	0.016
<i>CNTN4</i>	13	9	1.325	0.528	3.374	0.334
<i>CNTN5</i>	17	23	0.671	0.340	1.315	0.139
<i>CNTN6</i>	16	17	0.859	0.410	1.795	0.397
<i>CNTNAP1</i>	17	11	1.420	0.627	3.255	0.238
<i>CNTNAP2</i>	19	26	0.662	0.349	1.249	0.113
<i>CNTNAP4</i>	24	13	1.702	0.827	3.557	0.082
<i>CNTNAP5</i>	14	16	0.797	0.366	1.731	0.333

*rare variants were defined as follows: seen only in either cases or controls exclusively, missense, nonsense, splice site, or start or stop codon disruptions with a frequency of less than 2% in this data, and less than 1% in all populations in the Exome Variant Server [38] and SeattleSNP [39] databases

doi:10.1371/journal.pgen.1004852.t001

We reasoned that the background rate of neutral rare variation in the genome might be obscuring a true difference in functional alleles in cases versus controls. Consequently, we used a variety of defining criteria to try to distinguish the subset of potentially disease-relevant mutations: First, we evaluated “singleton” mutations, *i.e.* those seen only once either in cases or in controls. This strategy was intended to focus on the rarest alleles in our sample, those most likely to be subject to purifying selection. A comparison of rates of singleton mutations in *CNTNAP2* and all other sequenced *CNTN* and *CNTNAP* genes again found no evidence of association, prior to correction for the 10 genes evaluated in this study (Table 2). We then used two well-established informatics tools designed to predict deleterious amino acid substitutions. PolyPhen2 [21] and SIFT [19], [20]. We defined mutations as deleterious if they were either ‘possibly damaging’ or ‘probably

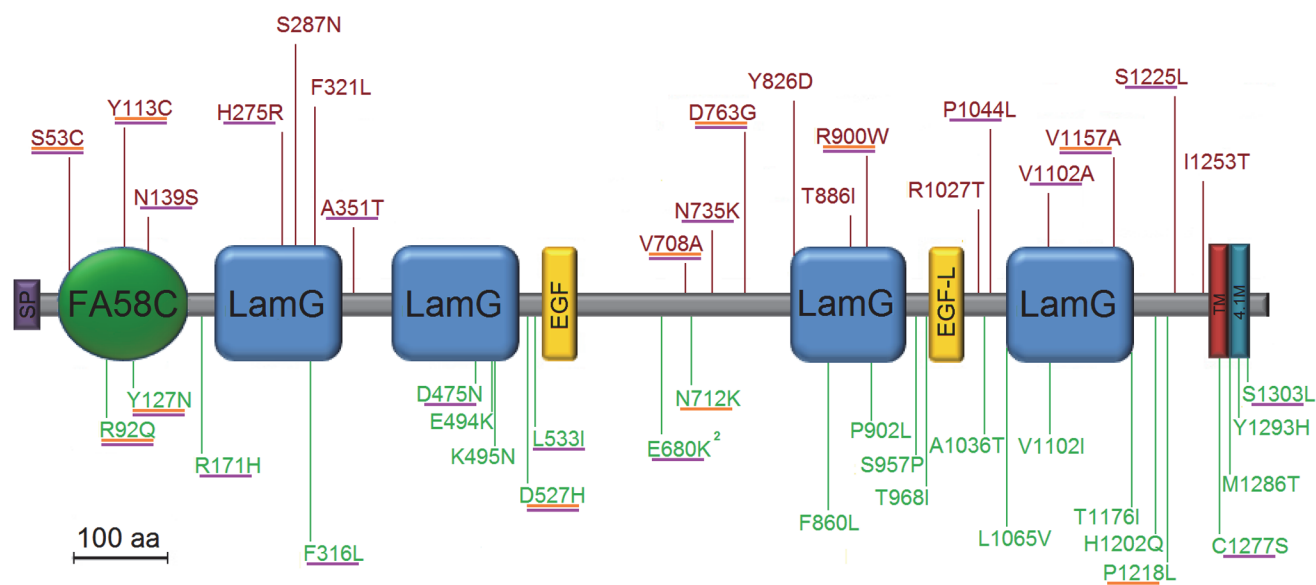


Figure 1. Location of all mutations of interest, *i.e.* rare and exclusive to cases or controls. Mutations were counted if missense, nonsense, splice site, or frameshift in *CNTNAP2* (in point of fact, all were missense). Variants in red are exclusive to cases; those in green, controls. Those predicted to be deleterious in SIFT are underlined in orange; in PolyPhen2, purple. Domain names and approximate locations from Bakkaloglu et al.,

doi:10.1371/journal.pgen.1004852.g001

damaging' in PolyPhen2 or 'damaging' in SIFT. While neither metric evaluates nonsense mutations and splice site disruptions, these were included as deleterious as well.

Using all of these approaches separately, we compared the rates of putatively deleterious/functional alleles in the same 1030 cases versus 942 controls. We found no significant differences using any of these metrics with regard to *CNTNAP2* (13 in cases, 13 in controls, Fisher exact one-tailed $p = 0.486$) (Table 3). *CNTN1* was found to show nominally significant association (12 in cases, 2 in controls, Fisher exact one-tailed $p = 0.01$) but this was not significant after correcting for the 10 genes studied. No other genes were significant prior to correction. Additionally, in an analysis restricted to singletons predicted to be deleterious by SIFT or PolyPhen2 (S1 Table), none of the genes reached significance prior to correction.

Finally, we restricted our analysis to focus solely on nonsense, splice site or frameshift mutations. We found none present in *CNTNAP2* in the current cohort. Moreover, this category of variation was so infrequent in both cases and controls in all other genes investigated that gene-by-gene analysis was not statistically meaningful, this limiting our conclusions about the role for true haploinsufficiency in any of the genes under study. Overall, we found 2 case individuals with a stop codon in one of the 10 genes (*CNTN2* and *CNTNAP1*), 1 case and no controls with a splice site mutation (*CNTN3*), and 3 control individuals displaying a stop codon (*CNTNAP1* and two in *CNTN6*). Of note, indel mutations were not detectable using the Raindance instrument given the short read length and the nature of the alignment process.

Of note, our primary analysis was restricted to only rare variants seen in cases or controls and not both. To be confident that this approach did not obscure a true difference in mutation burden, we re-ran all of the aforementioned analyses while including all point mutations defined as rare, regardless of whether they were present in both cases and controls. This had no impact on the interpretation of our results (S2-S3 Tables).

Power of burden test

For the primary analysis of *CNTNAP2*, we compute good power ($>80\%$) under the following parameters: observed rare variant frequency of 0.023 (Table 1), odds ratio for dominant effect of ≥ 2.0 , size of the test set to 0.005 (corrected for 10 tests), and for this sample of cases ($N = 1030$) and controls ($N = 942$). The empirically derived rare variant frequency for the average gene tested is lower (across all *CNTNs* and *CNTNAPs* including *CNTNAP2*) at 0.015. For such a gene, good power accrues for odds ratio ≥ 2.5 . Finally, if we consider the burden over all tested genes combined, for which the average frequency of rare variants is roughly 0.15, good power accrues if the odds ratio is ≥ 1.35 [22].

Analysis of I869T

In our 2008 report, we regarded I869T as a significant finding given its occurrence 3 times in cases and not at all in controls (Fisher exact one-tailed $p = 0.014$), its deleterious rating in SIFT, and its amino acid conservation [leucine in amphibians and fish, isoleucine in mammals]. Subsequent revisions to SIFT, however, have reclassified the mutation as benign, and its PolyPhen2 rating has remained benign. Furthermore, the expanded omnibus data set of 2704 cases and 2747 controls (S4 Table) revealed just one further occurrence in cases (in the Broad-Baylor cohort), none in controls; this was no longer statistically significant in the expanded sample size (Fisher exact $p = 0.06$) without correction for multiple comparisons.

De Novo mutation burden

In keeping with our recent findings [23] that the identification of recurrent *de novo* loss-of-function point mutations is a viable strategy to identify ASD genes, we determined the

Table 2. Rates of singleton* mutations: all genes.

Gene	Number in cases (n = 1030)	Number in controls (n = 942)	OR	OR 95% CI lower	OR 95% CI upper	One-tailed Fisher exact p-value
<i>CNTN1</i>	11	5	2.023	0.648	6.693	0.141
<i>CNTN2</i>	12	10	1.099	0.442	2.752	0.5
<i>CNTN3</i>	8	3	2.450	0.594	11.662	0.144
<i>CNTN4</i>	11	9	1.119	0.429	2.943	0.492
<i>CNTN5</i>	12	19	0.573	0.260	1.246	0.0900
<i>CNTN6</i>	16	17	0.859	0.410	1.795	0.397
<i>CNTNAP1</i>	15	11	1.251	0.540	2.927	0.359
<i>CNTNAP2</i>	19	24	0.719	0.375	1.372	0.180
<i>CNTNAP4</i>	20	11	1.676	0.761	3.750	0.115
<i>CNTNAP5</i>	14	16	0.797	0.366	1.731	0.333

*singleton mutations met the following criteria: seen only once in either cases or controls exclusively, missense, nonsense, splice site, or start or stop codon disruptions with a frequency of less than 2% in this data, and less than 1% in all populations in the Exome Variant Server [38] and SeattleSNP [39] databases

doi:10.1371/journal.pgen.1004852.t002

transmission status of every rare/singleton mutation seen in probands. We found 2 confirmed *de novo* mutations using whole blood DNA for sequencing reactions. No confirmed *de novo* mutations were observed in *CNTNAP2*. *CNTN6* and *CNTNAP4* each showed one *de novo* missense variant each. Based on recent analyses of exome data, a single missense *de novo* mutation, regardless of other characteristics such as degree of evolutionary conservation, does not provide significant evidence of association. [23–26]

Compound heterozygous mutations

We further examined individuals for the presence of multiple rare variants in the same gene: We found 2 probands with rare mutations, one in *CNTN6* and the other in *CNTNAP4*. A total of four controls were also identified with two rare mutations each: three were found in *CNTN5*, and one had two rare variants in *CNTN6*. We determined chromosome phase for the two cases and both were found to be compound heterozygotes. A similar analysis was not possible for

Table 3. Rates of mutation predicted deleterious* by SIFT-or-PolyPhen2: all genes.

Gene	Number in cases (n = 1030)	Number in controls (n = 942)	OR	OR 95% CI lower	OR 95% CI upper	One-tailed Fisher exact p-value
<i>CNTN1</i>	12	2	5.554	1.177	35.912	0.01
<i>CNTN2</i>	12	9	1.222	0.479	3.158	0.409
<i>CNTN3</i>	3	2	1.373	0.187	11.728	0.542
<i>CNTN4</i>	9	7	1.177	0.401	3.515	0.473
<i>CNTN5</i>	14	21	0.604	0.290	1.250	0.098
<i>CNTN6</i>	11	12	0.837	0.343	2.036	0.414
<i>CNTNAP1</i>	11	10	1.006	0.396	2.566	0.582
<i>CNTNAP2</i>	13	13	0.913	0.396	2.105	0.486
<i>CNTNAP4</i>	14	7	1.841	0.693	5.051	0.133
<i>CNTNAP5</i>	6	12	0.454	0.151	1.306	0.084

*deleterious mutations were defined as “damaging” in SIFT and/or “possibly damaging” or “probably damaging” in PolyPhen2, as well as all nonsense and splice site mutations which were not evaluated by these programs but were regarded as deleterious intrinsically

doi:10.1371/journal.pgen.1004852.t003

Table 4. Inheritance of mutations predicted deleterious* by SIFT-or-PolyPhen2.

Gene	Number paternal	Number maternal	One-tailed binomial p-value
<i>CNTN1</i>	9	3	0.073
<i>CNTN2</i>	4	8	0.1938
<i>CNTN3</i>	2	1	0.5
<i>CNTN4</i>	5	4	0.5
<i>CNTN5</i>	6	8	0.3953
<i>CNTN6</i> [†]	4	6	0.377
<i>CNTNAP1</i>	6	5	0.5
<i>CNTNAP2</i>	4	9	0.1334
<i>CNTNAP4</i> [†]	5	8	0.2906
<i>CNTNAP5</i>	4	2	0.3438

*deleterious mutations were defined as “damaging” in SIFT and/or “possibly damaging” or “probably damaging” in PolyPhen2, as well as all nonsense and splice site mutations which were not evaluated by these programs but were regarded as deleterious intrinsically. Inheritance was a situation where recurrent mutations were counted more than once, because keeping only one instance of a variant and ignoring the rest would introduce arbitrary parent-of-origin bias

[†]*CNTN6* and *CNTNAP4* each had one mutation that was confirmed to be de novo in whole blood

doi:10.1371/journal.pgen.1004852.t004

controls given the absence of parental data. Again, based on recent studies of whole-exome trio data in autism, the observation of a lone missense compound heterozygous mutation in a gene carries minimal statistical evidence for association [27].

Parental origin of mutation

With the family analysis data, we were able to determine inheritance status (S5 Table) for putatively deleterious mutations of interest in each gene. There was no significant difference in the proportion of deleterious (based on ratings in SIFT (damaging) and/or PolyPhen2 (possibly or probably damaging) variants maternally inherited and paternally inherited in any gene (Table 4). We also observed no significant difference between maternal and paternal variants when this gene-by-gene analysis was restricted to singletons (S6 Table).

Combined analysis of multiple cohorts for mutations in *CNTNAP2*

Given the divergent findings in our initial study [2] versus the new discovery cohort with regard to *CNTNAP2*, we expanded our dataset. Using exome sequencing data in both cases and controls, we were able to increase the total sample for this gene to 2704 cases and 2747 controls. Consistent with the findings above, an omnibus analysis in these 2704 cases and 2747 controls showed no significant results either in a primary analysis of rare mutation burden or in any of the exploratory analyses of variant subsets, as described above (rare mutation burden 45 in cases, 49 in controls, one-tailed Fisher exact $p = 0.407$, deleterious mutation burden 26 in cases, 28 in controls, one-tailed Fisher exact $p = 0.469$, deleterious singleton mutation burden 22 in cases, 23 in controls, one-tailed Fisher exact $p = 0.521$) (Tables 5, S7 and S8). Furthermore, the variants from our 2008 paper show an even weaker trend for overrepresentation of deleterious mutations in cases, due to updates to the SIFT and PolyPhen2 databases (see Tables 5, S7 and S8).

Discussion

Strauss et al.[1] first demonstrated a role for *CNTNAP2* in neurodevelopmental disorders based on a consanguineous pedigree in which probands presented with autism, cortical

Table 5. Combined *CNTNAP2* deleterious variants.**

	2008 Cohort * (635 cases and 942 controls)	Discovery cohort (1030 cases and 942 controls)	Baylor-Broad (1039 cases and 863 controls)	Cumulative (2704 cases and 2747 controls)
Case variants	(6)5 [†]	(13)12 [†]	9	26
Control variants	(9)8 [†]	13	(8)7 [†]	28
One-tailed Fisher exact p-value	.567	0.410	0.551	0.469

*reported in Bakkaloglu et al.[2]

**deleterious mutations were defined as “damaging” in SIFT and/or “possibly damaging” or “probably damaging” in PolyPhen2, as well as all nonsense and splice site mutations which were not evaluated by these programs but were regarded as deleterious intrinsically

[†]one variant was seen in a control from 2008 and case in this discovery cohort and was thus removed from this omnibus analysis (V1102A—benign in SIFT, possibly damaging in PolyPhen2); another variant in a 2008 case and a Baylor-Broad control (Y716C—benign in SIFT, possibly damaging in PolyPhen2) was also removed from this analysis

doi:10.1371/journal.pgen.1004852.t005

dysplasia and focal epilepsy. A single-base homozygous deletion resulting in a premature stop codon was identified in *CNTNAP2*[1]. In 2008, four studies provided varying degrees of evidence for a role for heterozygous mutations in *CNTNAP2* in non-syndromic autism and language delay. Bakkaloglu et al.[2] mapped a *de novo* chromosomal inversion disrupting *CNTNAP2* and then re-sequenced *the gene* in 635 cases and 942 controls, finding nominally significant association of a single rare variant and a trend toward significance for the overall burden of rare deleterious and conserved mutations in cases versus controls. Alarcon et al. [3] utilized family-based association with age at first word in 172 autism trios to identify *CNTNAP2* along with three other genes; in a second-stage follow-up in 304 independent trios, only *CNTNAP2* continued to show association with age at first word. Furthermore, microarray expression analysis identified 12 genes significantly enriched in human language-related association cortex, one of which was *CNTNAP2*. Arking et al.[4] used a linkage approach and identified a common variant in *CNTNAP2*; a second-stage independent cohort of 1295 autism trios showed significant overtransmission of the T allele. Finally, Vernes et al.[6], investigating the transcription factor FOXP2, previously implicated in language disorders [28–30], determined that FOXP2 binds to and downregulates *CNTNAP2*. Additional case reports of speech delay and/or ASD patients demonstrating disrupted *CNTNAP2* have subsequently been identified [7, 31]. Most recently, Peñagarikano [5] reported a *CNTNAP2* KO mouse that showed deficiencies in all three ASD core behavioral domains, seizures, and hyperactivity, with demonstrable neuronal migration aberrations, abnormal cellular morphology and a decreased number of interneurons.

Additionally, disruptions in other members of the contactin family and contactin-associated protein-like family have also been associated with autistic phenotypes. Multiple case reports have implicated contactin 4 (*CNTN4*) in ASD and in developmental delay [17, 32], as did a CNV study [33] described above. Contactin-associated protein-like 5 has also been recently implicated in ASD by a case study [18].

However, subsequent to these reports, the evidence for association of heterozygous variants in *CNTNAP2* has been variable. As noted, multiple GWAS analyses [11–13] have not identified evidence of genome-wide significant association of SNPs mapping near *CNTNAP2*. In addition, to date, neither rare *de novo* [23–26] or compound heterozygous mutations [27, 34, 35] have implicated *CNTNAP2* or other Contactin or Contactin Associated Proteins in autism or related phenotypes. Finally, other *CNTN* genes have so far not been subjected to large-scale case control rare variant analyses.

Against this backdrop, we undertook a follow-up of our earlier rare variant case control analysis [2] focusing on *CNTNAP2*, and expanded the study to include other *CNTN* and *CNTNAP* genes. As described above, our results do not convincingly replicate association of the previously nominally significant variant, I869T, nor support the association of heterozygous rare mutations, either transmitted or *de novo*, in *CNTNAP2* with the ASD phenotype. Nor is there evidence for rare recessive risk alleles in this outbred cohort. We found similar results for all *CNTN* and *CNTNAP* genes tested here. As noted above, in addition to investigating overall mutation burden, we conducted exploratory analyses to look for any potential differences between cases and controls in subsets of patients; these included evaluating predicted deleterious mutation burden, looking for multiple mutations within a single gene or across the genes sequenced in this study (see [Protocol](#)), investigating parental origin of the variation and evaluating for *de novo* mutations. While the primary analyses were well-powered to identify odds ratios in the range of 1.35 to 2.5, the secondary analyses are truly exploratory and cannot be interpreted as excluding a role for *CNTN* or *CNTNAP* genes in ASD.

There are several potential explanations for our inability to extend our previous findings: first, the prior results with regard to overall mutation burden showed only a trend toward significance and the larger, better powered sample studied here may simply be clarifying the true relationship (or absence of risk). Alternatively, the paucity of loss of function mutations in the expanded sample, coupled with the difficulty in interpreting the functional consequences of transmitted missense variants without a biologically validated assay could now be leading to a false negative result with regard to haploinsufficiency at this locus—though our results do strongly suggest that the overall contribution to population risk of heterozygous mutations of all types in *CNTN* and *CNTNAP2* must be quite limited. With regard to the prior finding of association of the I869T variant in cases, we did not find evidence for population stratification playing a role, but neither did we add substantial evidence for association, again limiting the plausible effect size of this risk.

As noted above, the lack of a reliable functional assay allowing us to differentiate biologically meaningful from neutral variants is an important limitation of the study. It is possible that such an assay could reveal a distribution of variation consistent with an effect for heterozygous mutations in *CNTNAP2*—or any of the other genes we have studied. Similarly, it was notable that we did not observe any putative loss of function mutations (nonsense, frame-shift, splice site) in either cases or controls for *CNTNAP2*. Given the finding [1] of a severe neurodevelopmental syndrome resulting from homozygous loss of function mutations, the observation of deletion CNVs at this locus in affected individuals, and the absence of clear LOF mutations in our sample, it remains possible that very rare LOF heterozygous mutations in *CNTNAP2* carry some risk for ASD. Nonetheless, based on the current study, our results place limits on the degree to which, overall, rare heterozygous missense mutations in Contactin and Contactin Associated Proteins may play a role in the population risk for ASD.

They similarly suggest that it is not yet possible to determine the clinical significance of individual rare heterozygous mutations in *CNTNAP2*, identified either in a research cohort or in diagnostic assays (<http://www.athenadiagnostics.com/content/test-catalog/find-test/service-detail/q/id/1487>).

Materials and Methods

Subjects and experimental structure

For the targeted re-sequencing discovery cohort, 1332 ASD probands were obtained from the Simons Simplex Collection (SSC), a cohort focused on simplex autism families [36]. Only probands from the SSC were sequenced. We elected to compare this cohort to 1015 unrelated

controls, determined to be neurologically unaffected, drawn from the NINDS Caucasian controls repository (Coriell, Camden, NJ). As described above and in detail in [S1 Protocol](#), the final case-control cohort was filtered to 1030 ASD probands and 942 controls. For the follow-up examination of *CNTNAP2*, we obtained exome sequencing data from the ARRA Autism Sequencing Collaboration (AASC). AASC cases and controls were genotyped on the Illumina 550K Single BeadChip platform and filtered to exclude any individual with a genotyping call rate <95%, discrepant genotype data with regard to reported sex, Mendelian inconsistencies, or cryptic relatedness, resulting in a final sample of 1039 autism cases and 863 controls, all of whom were determined to be of European descent through ancestry matching [37] via PCA of genotyping data. Controls were screened for psychiatric conditions via questionnaire. Although AASC/ARRA variants were not confirmed via Sanger sequencing as the Yale cohort of 1030 cases and 942 controls were, we concluded that given the identical sequencing methods for cases and controls as well as the approximately equal sample sizes, any errors would be reasonably randomly distributed among cases and controls. Sample characteristics and diagnostic methodology for the AASC have been described previously.⁵ This study was approved by the Institutional Review Boards of both Yale University and the Broad Institute.

Library preparation

Primer emulsions representing 1152 separate amplicons were designed and synthesized by Raindance Technologies (Lexington, MA). Synthesis was infeasible for much of *CNTNAP3* due to highly repetitive regions; because 16 exons out of 24 were excluded, this gene was not sequenced and therefore does not contribute to any further analyses. A single exon was omitted from each of the genes *CNTN2* (out of 22 exons total) and *CNTNAP4* (25 total) each due to difficulty in primer synthesis. We reasoned that any loss of data should be equally distributed across cases and controls.

In the initial sequencing effort, all DNA used was derived from human lymphoblastoid cell line DNA. All confirmations, however, were conducted in DNA derived from blood (see below). Samples were quantified on a Biotek Synergy HT fluorometer (BioTek US, Winooski, VT). After quantification, DNA was pooled, 8 individuals at a time, by case/control status ([S1 Protocol](#)).

Pooled DNA samples were sheared on a Covaris S2 (Covaris, Inc., Woburn, MA) to an approximate size of 3 kb and combined with RainDance microemulsion PCR master mix prepared according to the protocol, then run with a standard protocol on the RDT1000 machine (Raindance Technologies, Lexington, MA) and subjected to conventional PCR on a thermal cycler.

After successful droplet merging and PCR, the samples were cleaned according to the RainDance protocol; aqueous PCR product phase was purified on Qiagen MinElute columns (Qiagen GmbH, Hilden, Germany) and run on an Agilent Bioanalyzer according to DNA1000 protocol. Product was then blunted and concatenated into longer DNA fragments, because the range of amplicon sizes in the microemulsion library makes sequencing uniform fragment lengths impossible without first concatenation and subsequent shearing.

Concatenated DNA was sheared on a Covaris S2 to a mean size of 200 bp. Products were processed for end repair, adenylation and adapter ligation according to modified Illumina protocol ([S1 Protocol](#)). These fragments were size selected on Invitrogen agarose e-gels (Invitrogen, Life Technologies, Carlsbad, CA); size-selected DNA fragments were enriched using Illumina primers InPE 1.0, InPE 2.0, and a selected Illumina barcode index out of 12 available. Barcode indices were selected as described in [S1 Protocol](#). PCR was performed according to standard Illumina multiplexing protocol and samples were cleaned up on MinElute columns;

this product was then re-purified on an e-gel to remove primer dimer and all non-product fragments from the amplified product. 1 μ L of this product was run on an Agilent 2100 Bioanalyzer according to DNA1000 protocol to quantitate final concentration.

Sequencing of the discovery cohort

Separate pools of 8 case and 8 control samples were mixed in a 1:1 equimolar ratio and sequenced on the Illumina Genome Analyzer IIx. Sequence reads were run through a data analysis pipeline on a high performance cluster (HPC) and aligned to the human genome (NCBI build 36, hg18) using the aligner BWA, followed by conversion to SAM format and generation of a Pileup file using SAMtools. The Pileup file was then analyzed and annotated to identify variants with a haploid coverage greater than 80X, minor allele frequency (MAF) greater than 3.5% and a PHRED-like score (generated from the quality scores in the initial reads) greater than 20. The MAF and Phred-like score cutoffs were determined empirically. Annotation was performed for all possible isoforms of each gene to ensure that any deleterious variants were detected.

Variant confirmations

Pools in which putative rare mutations of interest were identified underwent confirmation and deconvolution using Sanger sequencing. Mutations of interest were defined as missense, nonsense, splice site, or start or stop codon disruptions with a frequency of less than 2%, and less than 1% in all populations in the Exome Variant Server [38] and SeattleSNP [39] databases. Only a single region showed repeated PCR failures (chr11:99437155, *CNTN5*). This variant was excluded from any further analyses. The confirmation rate of the high-throughput sequence variant predictions of interest, as determined from all variants which successfully underwent PCR and Sanger sequencing, was 92%.

After successful Sanger confirmations, all variants identified only in cases were tested for *de novo* status. An initial round of re-sequencing of the proband, parents and any siblings was completed using lymphoblastoid DNA; if the mutation appeared to be *de novo*, a second round of sequencing in all family members was completed using whole blood DNA to rule out false positives resulting from EBV transformation. Family member DNA was not available for inheritance evaluation in the Coriell controls.

As part of our analytic strategy we planned to evaluate *de novo* mutations, and consequently we required a high level of certainty regarding the identity of any proband carrying a putative mutation. Consequently, we developed an in-house script to verify pool composition by comparing returned pooled next-generation sequence data to previously obtained genotyping data available on all SSC probands reported here [40]. 15 pools (9 cases and 6 control) demonstrated insoluble identity issues (ambiguity of constituent individuals) and were excluded from all analyses.

Sample quality control and adjustments for comparable ancestry

SNP Genotyping was performed on Illumina 1M arrays at the Yale Center for Genome Analysis using whole blood DNA (SSC samples) or lymphoblastoid cell line DNA (NINDS controls). Redundant samples were removed from both the case and control cohorts. After sample removal, 1266 cases and 1015 controls genotyped concurrent with being run on Raindance. Of these, PLINK [41] was used to remove those with missing genotype data or call rates <95%, to remove inconsistencies in reported sex, to detect and remove Mendelian inconsistencies, and to identify and remove cryptically related samples using an assessment of inheritance by descent up to and including 2° relatives—after all other filtering steps, one proband was excluded due to relatedness.

Principal components analysis was performed with the genotype data from these samples using Golden Helix SNP and Variation Suite v7.5.4 (Bozeman, MT, USA), evaluating 8210 tagging SNPs not found to be in high linkage disequilibrium, used in the same way as described in a recent paper [42]. Based on a scree plot, we plotted the eigenvalues of the first 3 principal components which accounted for 32%, 23% and 11% of variation. We defined the final NINDS control set of 942 as a Caucasian cohort and observed that a threshold of 5 inter-quartile ranges from the third quartile contained all NINDS controls; we used this criterion as a cutoff. 50 SSC probands lay outside this threshold at 6 IQR or greater and consequently were excluded from further analysis.

As noted, we removed from the analysis any variants with a greater than 1% frequency in either the Exome Variant Server [38] or SeattleSNP [39] databases.

Exome cohort

The Baylor-Broad ARRA set consisting of 1039 cases and 863 controls was used to follow-up on our targeted sequencing data given its comparable size and ancestry-matched selection of probands [37].

Statistical analyses

All statistical tests were one-tailed Fisher exact tests, with the exception of parent-of-origin analysis (see below). We used a Bonferroni correction by a factor of 10, based on the 10 genes investigated. Any variant that was seen in both cases and controls was excluded from the analysis. Because parent of origin was a binomial outcome, per-gene one-tailed binomial tests were performed. Finally, in the combined *CNTNAP2* dataset of our 2008 data, the current data, and the Baylor-Broad ARRA data, any variants seen in cases and controls were removed from the set (resulting in the loss of one variant from the cumulative set.)

Supporting Information

S1 Protocol. Details of DNA quantitation, shearing, Raindance and PCR conditions, Illumina sequencing preparation, post-sequencing bioinformatics methods, and variant counting.
(DOCX)

S1 Table. The deleterious singleton mutation burden as defined by a rating of “damaging” in SIFT and/or “possibly damaging” or “probably damaging” in PolyPhen2.
(DOCX)

S2 Table. Rare mutation burden including variants seen in both cases and controls as well as those exclusive to cases or controls.
(DOCX)

S3 Table. Rare deleterious mutation burden including variants seen in both cases and controls as well as those exclusive to cases or controls.
(DOCX)

S4 Table. The individual mutations in *CNTNAP2* in cases and controls in this discovery set, the 2008 data as reported in Bakaloglu et al., and the Baylor-Broad data.
(DOCX)

S5 Table. Parent of origin of all inherited mutations in all genes, with deleterious status indicated.
(DOCX)

S6 Table. The inheritance patterns at all genes of singleton mutations regarded as deleterious.

(DOCX)

S7 Table. Overall rare mutation burden in *CNTNAP2* in cases and controls in this discovery set, the 2008 data as reported in Bakkaloglu et al., and the Baylor-Broad data.

(DOCX)

S8 Table. Deleterious singleton mutation burden in *CNTNAP2* in cases and controls in this discovery set, the 2008 data as reported in Bakkaloglu et al., and the Baylor-Broad data.

(DOCX)

Acknowledgments

We are grateful for the contributions of the management and staff at the Yale Center for Genomic Analysis, particularly Shrikant Mane, Milind Mahajan, Irina Tikhonova, Joshua Lubner, Kira Fitzsimons, David Harrison, Anna Rogers, Nicole Pizzoferrato, and Vanessa Spotlow, and to Nicole Wright, Sindhuja Kammela, and Jaclyn Jacobi for technical support. We thank T. Brooks-Boone and M. Wojciechowski for their help in administering the project. We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Particular gratitude is owed Bernie Devlin for his input on statistical power calculations. J. Murdoch is also particularly grateful to his Ph.D. thesis committee, Drs. Kenneth Kidd and James Noonan, as well as Dr. State, whose insight and guidance were invaluable to this project.

Author Contributions

Conceived and designed the experiments: ARG SJS MTM JDM MWS. Performed the experiments: JDM ARG MFW SA GTO MJR NMD NV ZW CS LG SYC. Analyzed the data: JDM ARG SJS JK MFW TVF MTM GTO MJR NMD ZW CS AJW BMN MJD. Contributed reagents/materials/analysis tools: BMN MJD. Wrote the paper: JDM MWS.

References

1. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, et al. (2006) Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N Engl J Med* 354: 1370–1377. doi: [10.1056/NEJMoa052773](https://doi.org/10.1056/NEJMoa052773) PMID: [16571880](https://pubmed.ncbi.nlm.nih.gov/16571880/)
2. Bakkaloglu B, O’Roak BJ, Louvi A, Gupta AR, Abelson JF, et al. (2008) Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am J Hum Genet* 82: 165–173. doi: [10.1016/j.ajhg.2007.09.017](https://doi.org/10.1016/j.ajhg.2007.09.017) PMID: [18179895](https://pubmed.ncbi.nlm.nih.gov/18179895/)
3. Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, et al. (2008) Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am J Hum Genet* 82: 150–159. doi: [10.1016/j.ajhg.2007.09.005](https://doi.org/10.1016/j.ajhg.2007.09.005) PMID: [18179893](https://pubmed.ncbi.nlm.nih.gov/18179893/)
4. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, et al. (2008) A common genetic variant in the neurexin superfamily member *CNTNAP2* increases familial risk of autism. *Am J Hum Genet* 82: 160–164. doi: [10.1016/j.ajhg.2007.09.015](https://doi.org/10.1016/j.ajhg.2007.09.015) PMID: [18179894](https://pubmed.ncbi.nlm.nih.gov/18179894/)
5. Penagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, et al. (2011) Absence of *CNTNAP2* leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147: 235–246. doi: [10.1016/j.cell.2011.08.040](https://doi.org/10.1016/j.cell.2011.08.040) PMID: [21962519](https://pubmed.ncbi.nlm.nih.gov/21962519/)

6. Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, et al. (2008) A functional genetic link between distinct developmental language disorders. *N Engl J Med* 359: 2337–2345. doi: [10.1056/NEJMoa0802828](https://doi.org/10.1056/NEJMoa0802828) PMID: [18987363](https://pubmed.ncbi.nlm.nih.gov/18987363/)
7. Poot M, Beyer V, Schwaab I, Damatova N, Van't Slot R, et al. (2010) Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder. *Neurogenetics* 11: 81–89. doi: [10.1007/s10048-009-0205-1](https://doi.org/10.1007/s10048-009-0205-1) PMID: [19582487](https://pubmed.ncbi.nlm.nih.gov/19582487/)
8. Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, et al. (2009) CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. *Am J Hum Genet* 85: 655–666. doi: [10.1016/j.ajhg.2009.10.004](https://doi.org/10.1016/j.ajhg.2009.10.004) PMID: [19896112](https://pubmed.ncbi.nlm.nih.gov/19896112/)
9. Zweier C (2012) Severe Intellectual Disability Associated with Recessive Defects in CNTNAP2 and NRXN1. *Mol Syndromol* 2: 181–185. doi: [10.1159/000331270](https://doi.org/10.1159/000331270) PMID: [22670139](https://pubmed.ncbi.nlm.nih.gov/22670139/)
10. Jackman C, Horn ND, Molleston JP, Sokol DK (2009) Gene associated with seizures, autism, and hepatomegaly in an Amish girl. *Pediatr Neurol* 40: 310–313. doi: [10.1016/j.pediatrneurol.2008.10.013](https://doi.org/10.1016/j.pediatrneurol.2008.10.013) PMID: [19302947](https://pubmed.ncbi.nlm.nih.gov/19302947/)
11. Anney R, Klei L, Pinto D, Regan R, Conroy J, et al. (2010) A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 19: 4072–4082. doi: [10.1093/hmg/ddq307](https://doi.org/10.1093/hmg/ddq307) PMID: [20663923](https://pubmed.ncbi.nlm.nih.gov/20663923/)
12. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, et al. (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459: 528–533. doi: [10.1038/nature07999](https://doi.org/10.1038/nature07999) PMID: [19404256](https://pubmed.ncbi.nlm.nih.gov/19404256/)
13. Weiss LA, Arking DE, Daly MJ, Chakravarti A (2009) A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461: 802–808. doi: [10.1038/nature08490](https://doi.org/10.1038/nature08490) PMID: [19812673](https://pubmed.ncbi.nlm.nih.gov/19812673/)
14. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, et al. (2012) Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* 21: 4781–4792. doi: [10.1093/hmg/dds301](https://doi.org/10.1093/hmg/dds301) PMID: [22843504](https://pubmed.ncbi.nlm.nih.gov/22843504/)
15. Sampath S, Bhat S, Gupta S, O'Connor A, West AB, et al. (2013) Defining the contribution of CNTNAP2 to autism susceptibility. *PLoS One* 8: e77906. doi: [10.1371/journal.pone.0077906](https://doi.org/10.1371/journal.pone.0077906) PMID: [24147096](https://pubmed.ncbi.nlm.nih.gov/24147096/)
16. Compton AG, Albrecht DE, Seto JT, Cooper ST, Ilkovski B, et al. (2008) Mutations in contactin-1, a neural adhesion and neuromuscular junction protein, cause a familial form of lethal congenital myopathy. *Am J Hum Genet* 83: 714–724. doi: [10.1016/j.ajhg.2008.10.022](https://doi.org/10.1016/j.ajhg.2008.10.022) PMID: [19026398](https://pubmed.ncbi.nlm.nih.gov/19026398/)
17. Fernandez T, Morgan T, Davis N, Klin A, Morris A, et al. (2004) Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. *Am J Hum Genet* 74: 1286–1293. doi: [10.1086/421474](https://doi.org/10.1086/421474) PMID: [15106122](https://pubmed.ncbi.nlm.nih.gov/15106122/)
18. Pagnamenta AT, Bacchelli E, de Jonge MV, Mirza G, Scerri TS, et al. (2010) Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biol Psychiatry* 68: 320–328. doi: [10.1016/j.biopsych.2010.02.002](https://doi.org/10.1016/j.biopsych.2010.02.002) PMID: [20346443](https://pubmed.ncbi.nlm.nih.gov/20346443/)
19. Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* 11: 863–874. doi: [10.1101/gr.176601](https://doi.org/10.1101/gr.176601) PMID: [11337480](https://pubmed.ncbi.nlm.nih.gov/11337480/)
20. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31: 3812–3814. doi: [10.1093/nar/gkg509](https://doi.org/10.1093/nar/gkg509) PMID: [12824425](https://pubmed.ncbi.nlm.nih.gov/12824425/)
21. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7: 248–249. doi: [10.1038/nmeth0410-248](https://doi.org/10.1038/nmeth0410-248) PMID: [20354512](https://pubmed.ncbi.nlm.nih.gov/20354512/)
22. Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19: 149–150. doi: [10.1093/bioinformatics/19.1.149](https://doi.org/10.1093/bioinformatics/19.1.149) PMID: [12499305](https://pubmed.ncbi.nlm.nih.gov/12499305/)
23. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, et al. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485: 237–241. doi: [10.1038/nature10945](https://doi.org/10.1038/nature10945) PMID: [22495306](https://pubmed.ncbi.nlm.nih.gov/22495306/)
24. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, et al. (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485: 242–245. doi: [10.1038/nature11011](https://doi.org/10.1038/nature11011) PMID: [22495311](https://pubmed.ncbi.nlm.nih.gov/22495311/)
25. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485: 246–250. doi: [10.1038/nature10989](https://doi.org/10.1038/nature10989) PMID: [22495309](https://pubmed.ncbi.nlm.nih.gov/22495309/)
26. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, et al. (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron* 74: 285–299. doi: [10.1016/j.neuron.2012.04.009](https://doi.org/10.1016/j.neuron.2012.04.009) PMID: [22542183](https://pubmed.ncbi.nlm.nih.gov/22542183/)

27. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, et al. (2013) Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 77: 235–242. doi: [10.1016/j.neuron.2012.12.029](https://doi.org/10.1016/j.neuron.2012.12.029) PMID: [23352160](https://pubmed.ncbi.nlm.nih.gov/23352160/)
28. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413: 519–523. doi: [10.1038/35097076](https://doi.org/10.1038/35097076) PMID: [11586359](https://pubmed.ncbi.nlm.nih.gov/11586359/)
29. MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, et al. (2005) Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am J Hum Genet* 76: 1074–1080. doi: [10.1086/430841](https://doi.org/10.1086/430841) PMID: [15877281](https://pubmed.ncbi.nlm.nih.gov/15877281/)
30. Zeesman S, Nowaczyk MJ, Teshima I, Roberts W, Cardy JO, et al. (2006) Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. *Am J Med Genet A* 140: 509–514. doi: [10.1002/ajmg.a.31110](https://doi.org/10.1002/ajmg.a.31110) PMID: [16470794](https://pubmed.ncbi.nlm.nih.gov/16470794/)
31. Petrin AL, Giacheti CM, Maximino LP, Abramides DV, Zanchetta S, et al. (2010) Identification of a microdeletion at the 7q33-q35 disrupting the CNTNAP2 gene in a Brazilian stuttering case. *Am J Med Genet A* 152A: 3164–3172. doi: [10.1002/ajmg.a.33749](https://doi.org/10.1002/ajmg.a.33749) PMID: [21108403](https://pubmed.ncbi.nlm.nih.gov/21108403/)
32. Roohi J, Montagna C, Tegay DH, Palmer LE, DeVincent C, et al. (2009) Disruption of contactin 4 in three subjects with autism spectrum disorder. *J Med Genet* 46: 176–182. doi: [10.1136/jmg.2008.057505](https://doi.org/10.1136/jmg.2008.057505) PMID: [18349135](https://pubmed.ncbi.nlm.nih.gov/18349135/)
33. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, et al. (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459: 569–573. doi: [10.1038/nature07953](https://doi.org/10.1038/nature07953) PMID: [19404257](https://pubmed.ncbi.nlm.nih.gov/19404257/)
34. Chahrouh MH, Yu TW, Lim ET, Ataman B, Coulter ME, et al. (2012) Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet* 8: e1002635. doi: [10.1371/journal.pgen.1002635](https://doi.org/10.1371/journal.pgen.1002635) PMID: [22511880](https://pubmed.ncbi.nlm.nih.gov/22511880/)
35. Yu TW, Chahrouh MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, et al. (2013) Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 77: 259–273. doi: [10.1016/j.neuron.2012.11.002](https://doi.org/10.1016/j.neuron.2012.11.002) PMID: [23352163](https://pubmed.ncbi.nlm.nih.gov/23352163/)
36. Fischbach GD, Lord C (2010) The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* 68: 192–195. doi: [10.1016/j.neuron.2010.10.006](https://doi.org/10.1016/j.neuron.2010.10.006) PMID: [20955926](https://pubmed.ncbi.nlm.nih.gov/20955926/)
37. Liu L, Sabo A, Neale BM, Nagaswamy U, Stevens C, et al. (2013) Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS Genet* 9: e1003443. doi: [10.1371/journal.pgen.1003443](https://doi.org/10.1371/journal.pgen.1003443) PMID: [23593035](https://pubmed.ncbi.nlm.nih.gov/23593035/)
38. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: <http://evs.gs.washington.edu/EVS/>), [December 2012].
39. SeattleSNPs, NHLBI Program for Genomic Applications, SeattleSNPs, Seattle, WA (URL: <http://pga.gs.washington.edu>) [August 2012].
40. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70: 863–885. doi: [10.1016/j.neuron.2011.05.002](https://doi.org/10.1016/j.neuron.2011.05.002) PMID: [21658581](https://pubmed.ncbi.nlm.nih.gov/21658581/)
41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575. doi: [10.1086/519795](https://doi.org/10.1086/519795) PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)
42. Fernandez TV, Sanders SJ, Yurkiewicz IR, Ercan-Sencicek AG, Kim YS, et al. (2012) Rare copy number variants in tourette syndrome disrupt genes in histaminergic pathways and overlap with autism. *Biol Psychiatry* 71: 392–402. doi: [10.1016/j.biopsych.2011.09.034](https://doi.org/10.1016/j.biopsych.2011.09.034) PMID: [22169095](https://pubmed.ncbi.nlm.nih.gov/22169095/)